РАДІОТЕХНІКА ТА ТЕЛЕКОМУНІКАЦІЇ

UDC 621.396.1
DOI https://doi.org/10.32782/2663-5941/2025.5.1/07

Vetoshko I.P.
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"

EFFICIENT RESOURCE ALLOCATION IN 5G SA: IMPACT OF SCHEDULING ON GBR AND NON-GBR TRAFFIC

The rapid development of fifth-generation networks in the Standalone (5G SA) variant poses new challenges for ensuring quality of service (QoS) in mixed traffic conditions, in particular guaranteed (GBR) and non-guaranteed (Non-GBR) traffic. VoNR (Voice over New Radio) voice services are particularly sensitive to latency, jitter and packet loss parameters, requiring the implementation of more flexible resource management mechanisms. Traditional scheduling algorithms, such as Round Robin, focus on uniform resource allocation and do not take into account service class or QoS dynamics, leading to instability of voice streams in a variable radio environment.

The article proposes a simulation approach to analysing resource scheduling logic in a 5G SA network using two schemes as examples: Round Robin and QoS-aware Adaptive. The study covers time profiles of throughput, latency, and tail distributions (p95/p99 percentiles) for both traffic classes. It is shown that the adaptive approach stabilises voice traffic near the service benchmark, reduces delay tails and ensures SLA compliance for GBR while controlling temporary reductions in Non-GBR performance. Integral metrics, including Jain's fairness index and cumulative distribution functions (CDF) of throughput, were used to verify the results, confirming the effectiveness of QoS-oriented scheduling.

The results of the study demonstrate that the implementation of adaptive schedulers in 5G SA improves the resilience of VoNR to radio channel fluctuations and ensures predictable service quality without the need for rigid resource reservation. The findings can be used to improve traffic management policies in modern 5G networks.

Key words: *VoNR, 5G Standalone, QoS, GBR, Non-GBR, adaptive scheduling, Round Robin, Adaptive scheduling (QoS-aware), SLA, delay, jitter, packet loss, throughput, Jain's fairness index, CDF, simulation modelling.*

Formulation of the problem. The introduction of 5G Standalone (SA) networks creates conditions for the full use of a wide range of services, among which VoNR (Voice over New Radio) voice connections are of critical importance. Unlike transitional technologies, the implementation of VoNR in a standalone 5G architecture places all responsibility for ensuring quality of service (QoS) on radio access resource management mechanisms. This places new demands on schedulers, who must ensure low latency and stable QoS metrics (delay, jitter, loss), but also take into account the specifics of mixed GBR/Non-GBR traffic, which is characteristic of modern telecommunications systems.

Traditional scheduling schemes, such as Round Robin, focus on uniform resource allocation and do not take into account service class or historical QoS behaviour. As a result, during peak times, voice traffic encounters excessive delays and losses, which reduces call quality. At the same time, QoS-aware approaches require a balance between guaranteed services (GBR) and best-effort flows (Non-GBR) to avoid prolonged degradation of secondary traffic. In addition, the complexity of the task is due to the need to simultaneously fulfil several interrelated conditions: stable compliance with SLAs for voice flows, minimisation of delay and jitter tails, limitation of unproductive retransmissions (HARQ), and maintenance of a high fairness index in the distribution of resources between traffic classes. Existing approaches tend to focus either on the stability of a single service class or on overall spectral efficiency, ignoring the complex impact of mixed traffic.

Thus, there is a need to develop and research adaptive schedulers capable of taking into account service criticality, radio conditions dynamics, and QoS parameter history in real time. Solving this problem is a key step towards ensuring high-quality VoNR voice connections and efficient resource utilisation in 5G Standalone networks.

Analysis of recent research and publications. The issue of quality of service (QoS) in fifth-generation (5G) mobile networks, especially in Standalone architecture, has received considerable attention in recent studies. Works [1, 6, 7] highlight the features of VoNR voice service implementation and approaches to improving their quality in conditions of variable radio channel state. The authors focus on key QoS indicators – delay, jitter, packet loss – and emphasise the need for adaptive resource scheduling to maintain voice service stability.

An important place is occupied by 3GPP standards [3, 4, 8, 13], which define the architecture of the 5GS system, describe the principles of operation of NR/NG-RAN and IP Multimedia Subsystem (IMS), and regulate the mechanisms of multi-level QoS provision. These documents form the technical basis for the development and implementation of resource management algorithms in 5G networks. Works [9, 10] propose adaptive resource allocation schemes using CQI and MCS, which demonstrate effectiveness in reducing delays and increasing throughput. At the same time, [11, 12] present industry approaches to building QoS architecture for 5G services, with a focus on the integration of resource management mechanisms and standardised traffic profiles. Recent studies [2, 5, 14] emphasise that QoS-aware scheduling is critical for voice traffic under mixed load conditions (GBR/Non-GBR). They confirm the advantages of adaptive approaches over classical uniform distribution schemes, while noting the trade-offs between flow performance and resource allocation fairness.

In general, existing publications form a solid theoretical foundation, but most works either focus on general aspects of 5G architecture or are limited to analysing a single class of service. The issue of comprehensive comparison of scheduling logics in mixed GBR/Non-GBR load scenarios remains insufficiently researched, which is the subject of this article.

The purpose of this study is to investigate the impact of resource scheduling logic in a 5G Standalone network on the quality of service for mixed traffic with guaranteed (GBR) and non-guaranteed (Non-GBR) speeds. The main focus is on modelling and comparing Round Robin and QoS-aware Adaptive schemes in the context of maintaining service requirements for voice streams over IMS/VoNR. The work aims to analyse the time profiles of throughput, delay and QoS variability for different traffic classes in order to determine the ability of adaptive algorithms to provide SLAs for voice services without excessive degradation of Non-GBR traffic performance. This approach allows striking a balance between prioritising critical services and fair resource allocation, as well as forming practical conclusions for improving scheduling policies in real 5G SA networks.

Summary of the main research material.

Mixed GBR/Non-GBR load in 5G SA and scheduling logic (RR vs Adaptive). The study considers mixed GBR/Non-GBR load in 5G Standalone (NR, TDD) with voice service over IMS. Traffic is carried over QoS flows with 5QI, which determines the acceptable delay, priority and (for GBR) guaranteed speed; 5QI = 1 is used for speech, 5QI = 5 for IMS signalling, and associated media streams with lower requirements can have 5QI = 6-9 [1]. This service criticality ‘label’ passes through the protocol layers and directly influences the scheduler's decision on the order and amount of PRB allocation in each time interval. To reproduce key resource allocation modes, two access plane logics are compared – uniform Round Robin (RR) and QoS-aware Adaptive – for both traffic classes. In the time series of throughput for GBR, an explicit benchmark of 50 Mbit/s is set, which allows quantitative tracking of deviations for each scheme; a set of scenarios (RR/Adaptive × GBR/Non-GBR) demonstrates how bandwidth time profiles change under different priorities and loads [2]. In this study, the GBR bearer is modelled as a synthetic high-bitrate service (50 Mbit/s) to stress-test scheduling under mixed traffic. It is not an IMS/VoNR voice GBR bearer; voice-related implications are assessed via latency/jitter/loss metrics, not GBR throughput values.

Within the proposed framework, 5QI is treated as a dynamic urgency signal: as cumulative delay or loss approaches service thresholds, the GBR traffic class is granted an additional share of PRBs even under deteriorating channel conditions; when CQI is favourable, resources are returned to Non-GBR. This aligns with ‘history-aware’ scheduling, where practical adaptive schemes consider not only instantaneous SNR/CQI but also QoS trends and short-term forecasts, augmented by heuristic adjustments under favourable radio conditions and service-class controls. Empirically, the multi-level response – channel state → QoS trend → prioritisation/transmission parameters – reduces losses, and avoids unnecessary UE energy expenditure during high-SNR phases [2].

The throughput time series reflect the expected asymmetry: for GBR with Adaptive, the curve stays close to 50 Mbit/s with minor fluctuations, while RR systematically underperforms, averaging around 40–42 Mbit/s. For Non-GBR, on the contrary, Adaptive gives way to GBR during peak loads: the achievable level remains around ~40 Mbit/s, while Round Robin often drops below 30 Mbit/s and is characterised by higher variability (Fig. 1). Such correlations between GBR and Non-GBR curves are a direct consequence of QoS prioritisation in dynamic resource allocation [3].

The time-varying delay pattern is equally revealing: GBR Adaptive shows a steady approach to the benchmark of ≈10 ms, while RR generates an increased average delay (>15 ms). For Non-GBR, Adaptive allows the delay to be maintained at ≈15 ms, while Round Robin exceeds 25 ms under high load. The overall conclusion is that ensuring SLA for voice requires dynamic QoS-aware scheduling, while the 'blind' uniformity of RR proves unsuitable for both GBR (due to non-compliance with targets) and Non-GBR (due to stretched queues at peaks) [3]. The ≈10 ms target is interpreted as the NR interface (RAN one-way) delay rather than an end-to-end PDB.

From a service architecture perspective, this is consistent with the role of IMS/VoNR: maintaining low radio latency at the gNB level directly affects call setup/maintenance time and subjective quality; accordingly, the sensitivity of the voice 5QI stream to delay tails requires prioritisation in channel degradation phases, which is implemented by an adaptive scheme. At the same time, the 'price' of such protection is fixed: at peak load moments, Non-GBR inevitably receives a smaller share of PRB, which manifests itself in a deterioration of instantaneous performance. This compromise highlights the advisability of further improving adaptive rules, taking into account the short-term forecast of QoS/channel metrics [2].

Comparative analysis of throughput in RR and Adaptive scenarios. A comparison of throughput time series for four scenarios (RR/Adaptive × GBR/Non-GBR) reveals characteristic differences in service requirement fulfilment. For the GBR flow, adaptive logic stabilises the curve near the 50 Mbit/s benchmark, reducing the amplitude of fluctuations and the duration of deviations from the target level. For Round Robin, by contrast, there is a systematic underperformance with an average level of about 40–42 Mbit/s and more noticeable local dips (Fig. 1). This difference is a direct consequence of taking into account the class of service and queue state in the adaptive approach, which allows PRB to be reallocated in favour of GBR during critical intervals without rigid resource reservation [4].

For Non-GBR, the effect manifests itself as a 'controlled concession': during periods of increased load, the adaptive approach temporarily reduces the share of PRB in favour of GBR, which lowers the instantaneous throughput of the Non-GBR flow; outside of peaks, the resource is returned, and the Non-GBR curve stabilises at values higher than RR. In Round Robin, where the service class is ignored, both flows are in a common deficit – GBR does not reach the benchmark, and Non-GBR remains at even lower levels, with greater variability over time. In general, this corresponds to a negative correlation between the GBR and Non-GBR curves specifically for adaptive allocation, which reflects the prioritisation mechanism [2].

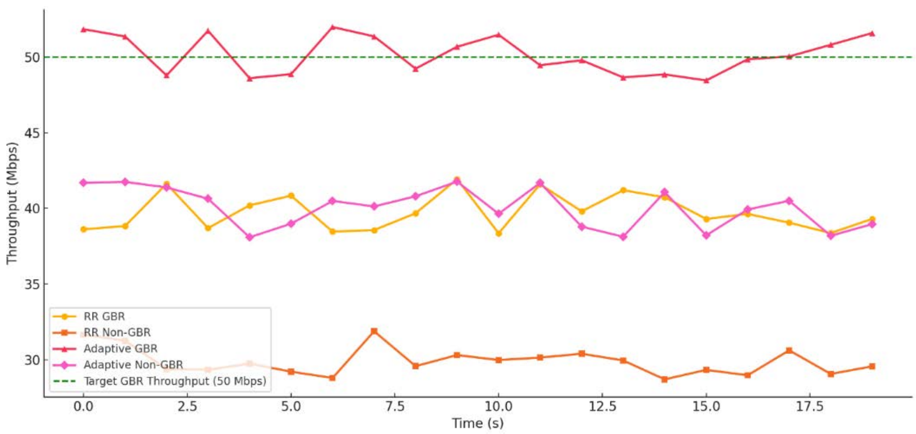

Fig. 1. Throughput comparison for GBR and Non-GBR traffic under different schedulers

From a scientific perspective, it is important to evaluate not only average values but also temporal characteristics of the throughput signals. At the trend level, the adaptive approach demonstrates less dispersion and a shorter transition process after local channel deteriorations: the return to the operating plateau occurs faster, which reduces the integral ‘cost’ of fluctuations for the voice stream. In terms of operational indicators, this is equivalent to a smaller proportion of time spent below the GBR benchmark and a lower coefficient of variation for Adaptive versus RR. Accordingly, for Non-GBR in adaptive mode, it is advisable to look not only at the averages, but also at the structure of the variability: interpeak intervals are characterised by higher predictability, while short load peaks contribute localised reductions without long ‘trails’. The formal recording of observations is reinforced by the presentation of cumulative distributions [5]. The CDF of throughput for GBR with adaptive distribution shifts to the right and fits more closely to 50 Mbit/s, clearly demonstrating a higher probability of reaching/staying near the benchmark; for Round Robin, the curve for GBR is shifted to the left, with a ‘heavier’ left wing. For Non-GBR, the difference in CDF highlights the ‘price of prioritisation’: under Adaptive, the probabilities of high values at peaks are lower, but between peaks the curves are more compact, reflecting the controllability of redistribution.

A useful addition to Fig. 1 is the integrated metrics on the summary panel: the proportion of time during which GBR is below 50 Mbit/s; p5/p50/p95 throughput for each ‘scheduler × class’ combination; Jain's fairness index between classes (by PRB share) as a quantitative assessment of the compromise between ‘GBR stability ↔ Non-GBR performance’ [6]. Taken together, these metrics reinforce the graphical picture and allow conclusions to be drawn that are reproducible regardless of the choice of observation points. From the perspective of radio access mechanisms, the advantage of the adaptive approach comes from combining service class information (via 5QI) and the current radio state in the form of CQI/SNR: at moments of channel deterioration for voice, PRB reallocation and transmission parameter correction (including MCS change) occur earlier and with less ‘inertia’ than in neutral schemes. This explains the shallower depth and shorter duration of dips for GBR and, accordingly, better resistance to short-term fluctuations [7]. For Non-GBR, the result is controlled degradation at peaks without the accumulation of long queues, which is important for maintaining the overall efficiency of the system.

Summarising the results, the time series of throughput and their derived indicators confirm three key conclusions: firstly, the adaptive approach consistently maintains GBR at a level close to 50 Mbit/s throughout the observation interval; second, Round Robin systematically fails to meet GBR requirements and at the same time does not provide favourable conditions for Non-GBR, leaving both flows in a common deficit; third, the ‘price’ of adaptive prioritisation for Non-GBR is short-term and manageable, which is reflected in more compact CDF curves and lower inter-peak variability.

Time profiles of delay, jitter, and loss; implications for VoNR/IMS. Time profiles of delay for mixed traffic show fundamentally different behaviour depending on the scheduler class. For GBR, in the case of adaptive allocation, the values consistently approach the benchmark of ≈10 ms, which corresponds to compliance with QoS requirements during most observation intervals; with Round Robin, the average delay exceeds 15 ms. For Non-GBR, the adaptive approach maintains a level of ~15 ms, while RR peaks exceed 25 ms (Fig. 2). These differences are consistent with time series data and directly indicate the advantage of QoS-aware scheduling for services with guaranteed characteristics. This can be explained by the mechanism of head-of-line delay accumulation in neutral allocation schemes: when each active UE receives approximately the same share of PRBs, voice packet queues grow even during short-term channel deteriorations [8]. The adaptive approach uses service criticality ‘tags’ (5QI), so at critical moments for GBR, prioritisation is performed with correction of the PRB share and transmission parameters; with favourable CQI, the resource is returned to Non-GBR. This logic reduces the duration and depth of deviations from target values and prevents long ‘trails’ of delays, while maintaining overall spectral efficiency [9].

To quantify the behaviour of the distribution tails, the p95/p99 percentiles and the proportion of intervals exceeding the threshold values relevant to each service class are used (Table 1). For GBR, the adaptive scheme demonstrates a significant compression of the tails: p95 = 12.5 ms and p99 = 14.0 ms versus 22.0 ms and 27.0 ms in RR; the proportion of intervals above 10 ms is 28% (Adaptive) versus 88% (RR). For Non-GBR, the adaptive approach keeps the median at 15.3 ms with p95 = 20.8 ms and p99 = 24.5 ms and only 6% of intervals exceeding 25 ms, while RR has a median of 22.7 ms, p95 = 31.5 ms, p99 = 39.0 ms, and 42% of threshold exceedances [5]. This ratio confirms that adaptive prioritisation not only reduces

Table 1

Delay distribution tails for scheduler–traffic-class combinations

Scheduler	Traffic class	Median, ms	p95, ms	p99, ms	Proportion > threshold, %	Threshold value, ms
RR	GBR	16.8	22.0	27.0	88	10
Adaptive	GBR	10.2	12.5	14.0	28	10
RR	Non-GBR	22.7	31.5	39.0	42	25
Adaptive	Non-GBR	15.3	20.8	24.5	6	25

the average values for GBR, but also systematically 'trims' critical tails, while keeping Non-GBR within acceptable limits outside of peaks.

The practical effect manifests itself at the IMS/VoNR procedure level: reducing radio delay on the NR interface shortens the RAN component of call-setup latency (INVITE→200 OK) and the time to first 180/183 (early-media onset). The obtained p95/p99 profiles and exceedance rates confirm that QoS-aware scheduling provides greater predictability for both guaranteed services (GBR) and best-effort traffic (Non-GBR) during off-peak intervals without requiring strict resource reservation [10].

The further relationship between RAN and service quality can be conveniently described by the integral 'threshold exceedance area' for voice: this refers to the total time and depth of delay exceedances beyond the 10 ms limit. A comparison with Table 1 shows that the adaptive approach not only reduces p95/p99 for GBR, but also significantly reduces the duration of such episodes. This correlates well with practical quality indicators (in particular, subjective assessment of voice service), since it is the 'tails' and long episodes of threshold exceedance that are most painful for the user [11]. Prioritisation based on adaptive logic works in short bursts: when the head-of-line delay increases or there is a noticeable deficit to the GBR target parameters, the scheduler temporarily increases the share of PRB for voice, and then restores the balance. This is confirmed by Table 2: the average PRB share is shifted in favour of GBR (≈58/42%), but the fairness index remains high (J≈0.975), i.e. the asymmetry is controlled and temporary. In practice, it is precisely this 'breathing' behaviour that allows GBR queues to be removed without prolonged depletion of Non-GBR.

Table 2

Average PRB allocation by traffic class and Jain's fairness index (RR vs Adaptive)

Scheduler	PRB for GBR, %	PRB for Non-GBR, %	Jain fairness Index (between classes)
RR	50.0	50.0	1.0000
Adaptive	58.0	42.0	0.9750

From an implementation standpoint, a two-loop design is advisable. The internal circuit monitors the 'quality deficit' for GBR and, if necessary, increases its priority; the external circuit maintains the framework – the minimum PRB share for Non-GBR at peaks and the upper limit of the instantaneous share for GBR – so that prioritisation does not 'drag on' after unloading. This combination explains why you have both low p95/p99 for GBR (Table 1) and a high fairness score (Table 2).

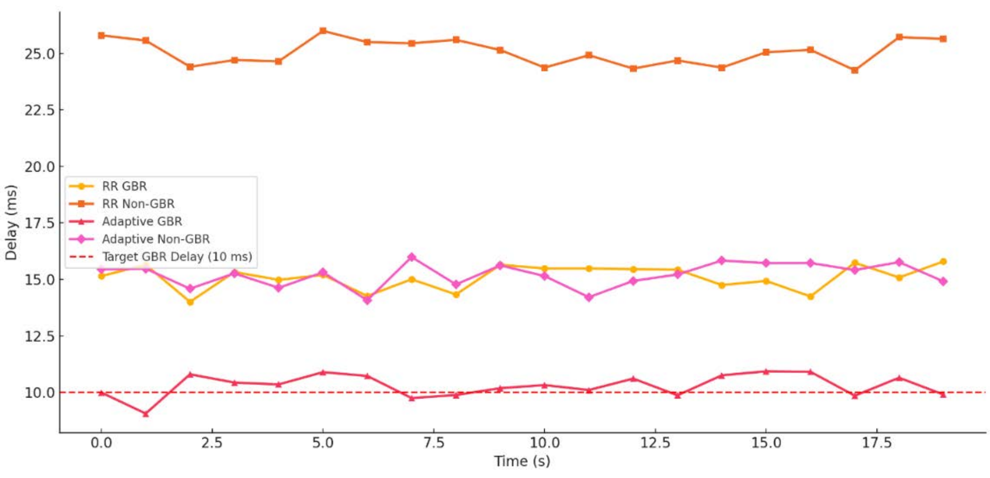

Fig. 2. Latency comparison for GBR and Non-GBR traffic under different schedulers

Sensitivity to the load profile is expected and useful for settings. As the share of voice sessions grows, it makes sense to slightly increase the prioritisation weight for GBR, but at the same time raise the minimum share for Non-GBR – so that J remains within an acceptable range. If best-effort dominates, the opposite is true: weaken the priority for GBR and more strictly limit the peak share for voice. In TDD profiles with longer downlink slots, the system better 'digests' short peaks – you can increase the smoothing windows; in more symmetrical or UL-heavy patterns, the response should be accelerated [13].

A useful operational metric is 'violation area' (total 'overshoot' of more than 10 ms for GBR). It summarises well the changes in p95/p99 and the percentage of overshoots in Table 1 and is directly related to unwanted retransmissions and MCS jitter. With adaptive allocation, this area is expected to be smaller, and therefore the probability of degradation that the user notices in real calls is also smaller. To transfer settings from the model to the network, it is worth establishing a simple order [12]. First, select the lowest prioritisation weight at which the target for GBR tails is met (e.g., p99 □ 14 ms and an acceptable proportion of time above 10 ms). Next, select the limits for the minimum Non-GBR share and the maximum instantaneous GBR share so that the fairness index does not fall below the selected threshold (e.g., 0.95) and the Non-GBR 'starvation' episodes are short. Finish by checking the stability of decisions for ±20% load variation and alternative TDD patterns. A separate practical advantage is terminal energy efficiency [14]. When delay tails are cut off and throughput is less 'jerky,' the number of retransmissions and aggressive MCS reconfigurations decreases. Although UE power consumption was not a target metric in this study, the observed stabilisation indirectly indicates no negative effects on the battery compared to neutral schemes [11].

The applicability of the results is determined by the model conditions: fixed traffic profiles, HARQ/RLC parameters, and radio channel stability within the series. In the presence of intense interference events or additional service classes, recalibration of weights and distribution limits will be required [11]. At the same time, the key pattern remains: short, controlled prioritisation pulses are sufficient to significantly reduce delay tails for GBR without causing prolonged degradation of Non-GBR, which is consistent at the level of two independent groups of indicators (Tables 1–2). Accordingly, the 50 Mbit/s GBR throughput target refers to the synthetic GBR service used for load stressing, not to VoNR/IMS voice. The ≈10 ms target is interpreted as the NR interface (RAN one-way) delay rather than an end-to-end PDB.

Conclusions. The analysis shows that QoS-aware adaptive scheduling under mixed GBR/Non-GBR traffic conditions in 5G SA ensures stable approximation of voice traffic to service targets. At the throughput level for GBR, the curve stays close to the 50 Mbit/s target, while Round Robin shows systematic underperformance; for Non-GBR, adaptive logic forms a controlled, short-term 'concession' of resources at peaks with balance restoration outside them (Fig. 1). Delay time profiles confirm the advantage of the adaptive approach for services with guaranteed characteristics. Significant compression of 'tails' is observed for GBR (p95/p99 and the proportion of exceedances of 10 ms) and moderate, localised degradation of Non-GBR at peaks compared to Round Robin (Fig. 2, Table 1). Such compression of the 'tails' directly reduces the 'threshold exceedance area' and correlates with an improvement in the perceived quality of speech in VoNR, as well as with a reduction in unproductive retransmissions.

PRB resource profiles are consistent with the quality indicators obtained: the average PRB allocation asymmetry in favour of GBR (≈58/42%) is accompanied by a high Jain's fairness index between classes (J≈0.975), reflecting the controlled nature of prioritisation and the absence of prolonged 'starvation' of Non-GBR traffic (Table 2). At the IMS/VoNR level, these PRB profiles align with the observed quality indicators and contribute to reducing the RAN component of call-setup latency (INVITE→200 OK) and the time to first 180/183 (early-media onset) [12].

It is important in practical terms that achieving the target indicators does not require strict resource reservation: short prioritisation pulses with restrictions on the minimum PRB share for Non-GBR and the upper instantaneous share for GBR are sufficient. The proposed dual-loop control logic (the inner loop responds to QoS deficits, while the outer loop maintains fairness) provides a reproducible compromise for different load profiles and TDD patterns. The limitations of the study relate to model assumptions (constant HARQ/RLC parameters, traffic profiles, radio channel scenarios) [11]. In the presence of intense interference events or when the proportion of service classes changes, it is advisable to recalibrate the prioritisation weights and distribution limits. At the same time, the basic pattern remains: short, bounded prioritisation pulses significantly reduce GBR delay tails without prolonged Non-GBR degradation, as confirmed by two independent sets of indicators (Tables 1–2).

Bibliography:

1. Ветошко І.П., Кравчук С.О. Структурні особливості реалізації голосових послуг VONR в мережі мобільного зв'язку стандарту 5G. *Вчені записки Таврійського національного університету імені В.І. Вернадського Серія: Технічні науки*. 2024. Том 35, No 1. P. 27–33.

2. Ветошко І.П., Кравчук С.О. Пріоритизація голосового трафіку в 5G: роль планувальників у забезпеченні QOS. *Дев'ятнадцята міжнародна науково-технічна конференція "Перспективи телекомунікацій"*, (Kyiv, 14-18 April 2025) Kyiv, 2025. P. 189–192.

3. 3GPP TS 23.501 version 18.5.0 Release 18. System architecture for the 5G System (5GS). Effective from 2025-03. Official edition. FRANCE: 650 Route des Lucioles F-06921 Sophia Antipolis Cedex, 2025. 542 p.

4. 3GPP TS 38.300 version 18.4.0 Release 18. NR and NG-RAN Overall Description; Stage 2. Effective from 2025-01. Official edition. FRANCE: 650 Route des Lucioles F-06921 Sophia Antipolis Cedex, 2025. 272 p.

5. Abuajwa O., Roslee M., Yusoff Z.B., Chuan L.L., Leong P.W. Resource Allocation for Throughput versus Fairness Trade-Offs under User Data Rate Fairness in NOMA Systems in 5G Networks. *Applied Sciences.* 2022. Vol. 12, No. 7. P. 3226. DOI: https://doi.org/10.3390/app12073226

6. Vetoshko I.P., Kravchuk S.O. Possibilities of improving the voice services quality in 5G networks. *Information and Telecommunication Sciences*. 2023. Vol.14, No 2. P. 9–16.

7. 1. Vetoshko I., Noskov V. Evaluation of voice transmission quality in the LTE networks. *Information and Telecommunication Sciences*. 2021. No. 2. P. 22–26.

8. 3GPP TS 26.114 version 18.3.0 Release 18. IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction. Effective from 2025-06. Official edition. FRANCE: 650 Route des Lucioles, F-06921 Sophia Antipolis Cedex, 2025. 517 p

9. Kwon J., Park J., Rhee W. A QoS-aware adaptive resource allocation scheme using CQI and MCS for LTE-A networks. *Wireless Personal Communications*. 2014. Vol. 76. P. 193–207.

10. Vora P. Voice over 5G (Vo5G): Technical Challenges and Performance Evaluation. *IEEE Communications Magazine*. 2021. Vol. 59, No. 10. P. 16–22.

11. Vetoshko I., Kravchuk S. Integration of machine learning-based prediction and dynamic QoS optimization into adaptive VoNR scheduling in 5G standalone networks: a simulation-based approach. *Вчені записки Таврійського національного університету імені В.І. Вернадського*. *Серія: Технічні науки*. 2025. Том 36 (75), No 3. P. 42–55.

12. Jain R., Chiu D.-M., Hawe W. A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems. – Research Report TR-301. Digital Equipment Corporation, Hudson, MA: Digital Equipment Corporation, 1984. URL: https://www.cse.wustl.edu/~jain/papers/ftp/fairness.pdf (дата звернення: 02.08.2025).

13. 3GPP TS 38.401 version 18.3.0 Release 18. NG-RAN; Architecture description. Effective from 2024-09. Official edition. FRANCE: 650 Route des Lucioles, F-06921 Sophia Antipolis Cedex, 2024. 161 p.

14. Vetoshko I.P., Kravchuk S.O. Opportunities to Improve the Quality of Voice Services in 5G Networks. *IEEE International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo)*. Kyiv, Ukraine. 2023. P. 92–97.

Ветошко І.П. ЕФЕКТИВНИЙ РОЗПОДІЛ РЕСУРСІВ У 5G SA: ВПЛИВ ПЛАНУВАННЯ НА ТРАФІК GBR І NON-GBR

Стрімкий розвиток мереж п'ятого покоління у варіанті Standalone (5G SA) формує нові виклики для забезпечення якості обслуговування (QoS) в умовах змішаного трафіку, зокрема гарантованого (GBR) та негарантованого (Non-GBR). Голосові сервіси VoNR (Voice over New Radio) є особливо чутливими до параметрів затримки, джитера та втрат пакетів, що вимагає впровадження більш гнучких механізмів управління ресурсами. Традиційні алгоритми планування, такі як Round Robin, орієнтовані на рівномірний розподіл ресурсів і не враховують клас сервісу чи динаміку QoS-показників, що призводить до нестабільності голосових потоків у змінному радіосередовищі.

У статті запропоновано симуляційний підхід до аналізу логік планування ресурсів у мережі 5G SA на прикладі двох схем: Round Robin та QoS-обізнаної Adaptive. Дослідження охоплює часові профілі пропускної здатності, затримки та «хвостів» розподілу (p95/p99) для обох класів трафіку. Показано, що адаптивний підхід стабілізує голосовий потік на рівні сервісного еталона, скорочує «хвости» затримки та забезпечує виконання SLA для GBR при контрольованому тимчасовому зниженні продуктивності Non-GBR. Для верифікації результатів застосовано інтегральні метрики, включно з індексом справедливості Джейна та кумулятивними функціями розподілу пропускної здатності, що підтверджує ефективність QoS-орієнтованого планування.

Результати дослідження демонструють, що впровадження адаптивних планувальників у 5G SA дозволяє підвищити стійкість VoNR до коливань радіоканалу та забезпечити передбачувану *якість сервісу без необхідності жорсткого резервування ресурсів. Отримані висновки можуть бути використані для практичного вдосконалення політик управління трафіком у сучасних мережах 5G.*

Ключові слова: *VoNR, 5G Standalone, QoS, GBR, Non-GBR, адаптивне планування, Round Robin, QoS-aware, SLA, затримка, jitter, втрати пакетів, пропускна здатність, індекс справедливості Джейна, CDF, симуляційне моделювання.*

Дата надходження статті: 25.09.2025
Дата прийняття статті: 16.10.2025
Опубліковано: 16.12.2025